An energy-based SOM model not requiring periodic boundary conditions

Alexander Gepperth
University of Applied Sciences Fulda
Department of Applied Computer Science
Leipzigerstr. 123, 36036 Fulda, Germany
Email: alexander.gepperth@cs.hs-fulda.de

***Abstract*—We present a SOM model based on a continuous energy function derived from the original energy-based model developed in [9]. Due to the convolution that is contained in the energy function, this model can only be applied when periodic boundary conditions are imposed (toroidal SOM), leading to markedly higher quantization errors, especially for small map sizes. We introduce a simple strategy, based on the assumption of homogeneous long-term averages for input-prototype distances, that allows to operate the energy-based SOM model without periodic boundary conditions, and demonstrate that its quantization errors are consistently lower especially for small map sizes. Simple experiments are conducted showing the worth of a continuous energy function, namely for novelty detection and automatic control of SOM parameters.**

I. Introduction

This article is in the context of self-organized map (SOM) models that have a continuous energy function. The lack of such an energy function for the original SOM model [11] has been the subject of a multitude of articles [4], [14], and several proposals were made to remedy this problem. The advantages of models whose learning rule is derived from the minimization of an energy function are numerous, while the only disadvantages are that the existence of an energy function imposes strong constraints on the used learning rules. In particular, it was shown that the original SOM learning rule cannot be derived from a continuous energy function[9]. In general, one may cite the following advantages of energy-based SOM models:

- **Estimation of learning success and parameter selection** A big issue for SOMs is to know whether the model has converged to a "desirable" state. For problems that do not allow a visual quality inspection of the learned prototypes (such as can be performed for the MNIST benchmark), there is no universal criterion to determine optimal values for the model parameters (final neighbourhood radius, final learning rate, topology etc.), whereas an energy function provides a scalar value that can be compared.
- **Proof of stability** If a continuous energy function exists and is bounded from below, this automatically guarantees the eventual convergence of SOM learning.
- **Use of advanced stochastic gradient descent methods** With a continuous energy function, many widely-used methods for performing stochastic gradient descent (SGD) in the domain of deep learning can be transferred to SOM learning. This is because methods such as RMSProp, Adam, AdaGrad, AdaDelta or "normal" SGD [6], which try to find "good" minima of an energy function, implicitly assume that such a quantity exists when adapting the meta-parameters of learning.
- **Outlier detection** When the energy (that is supposed to be minimized by the learning process) suddenly increases, this is a strong indication for a change in data statistics and thus be used for outlier detection or just to detect drift in the data. This latter property is especially relevant for our own ongoing work on incremental learning methods[5]. On the contrary, if there is no energy function, it is not even very clear what quantity can be used for outlier detection, as there is no function that is minimized by learning.

Fig. 1. Representative samples from the real-world pedestrian detection dataset used in this study. Incoming samples have to be classified into one of two classes "pedestrian" and "background".

A. Related work on energy-based SOM models

There has been a huge amount of primarily mathematical literature about It was shown conclusively in [4] that the original Kohonen learning rule cannot be exactly derived from the minimization of *any* error function. In the same article, it is mentioned that the Kohonen learning rule follows instead from the individual minimization of per-neuron energy functions [14], but these functions are very complex, non-unique and

978-1-5090-6638-4/17/$31.00

do not lend themselves to a simple interpretation (e.g., minimization of a distortion measure or similar). Another approach was proposed by Kohonen[11] and taken further by Heskes[9]: instead of finding error functions whose minimization would lead to the Kohonen learning rule, these authors attempted to very slightly modify the Kohonen rule such that an energy function could be found. Obviously, the modification should in no way impair the self-organization capabilities of the model while allowing an intuitive interpretation through a (preferably simple) energy function. An modification satisfying these requirements was proposed in [9], [8], offering a continuous energy function for discrete as well as continuous data distributions. While this was an important theoretical result, curiously enough there was no real follow-up in terms of applications in data visualization and/or clustering, which is surprising given the advantages an energy function can offer for SOM training, see above. It may be supposed that this lack of interest was due to the added computational complexity (an additional convolution needs to be calculated), as well as the problems that convolutions encounter at boundaries. Similar SOM variants having an energy function were proposed in [7] but they suffer from the same "convolution problem".

II. Methods and data

In all experiments, we use the energy-based SOM model as proposed by Heskes[9]. It is identical to the original SOM model except for the determination of the best-matching unit during a learning step, which involves a convolution of input-prototype distances. By virtue of this slight modification, it can be shown very easily that the model has a continuous energy function that is bounded from below.

A. Energy-based SOM model

In more precise terms, we consider quadratic two-dimensional maps M of $n \times n$ units, identified either by a linear index i or a coordinate tuple (i, j). Unless otherwise stated, we will use the linear index. Each unit with index i is associated with a prototype vector $\vec{p}_i$. The map receives, at each time step t, an input $\vec{x}(t)$ and transforms it into input-prototype distances $D = \{d_i(t)\}$ by taking the euclidean distance (dropping the dependency of all quantities on t for clarity):

$$d_i = ||\vec{p}_i - \vec{x}||$$

As usual for SOMs, we define local neighbourhoods by a Gaussian neighbourhood function $\gamma(\vec{\mu}, \sigma, \vec{x}) = \exp(-\frac{(\vec{x}-\vec{\mu})^2}{2\sigma^2})$ with a center $\vec{\mu}$ and a time-dependent variance $\sigma(t)$. However contrary to usual SOM practice, we express γ in a $n \times n$ *neighbourhood filter* $G = \{g_{ij}\}$ at each time step and use it in the determination of the best-matching unit (BMU). G is explicitly two-dimensional, centered on the map center and has entries given by

$$\vec{\mu} = (\frac{n}{2}, \frac{n}{2})^T$$
$$g_{ij} = \gamma(\vec{\mu}, \sigma, (i, j)^T)$$

In contrast to the original SOM model, the 2D indices (i^*, j^*) of the best-matching unit (BMU) for the energy-based Heskes model are determined by taking the 2D convolution

$$\begin{aligned} a_{ij} &= (D * G)_{ij} = \\ &= \sum_{i',j'=0}^{n} g_{i'j'} d_{i+\frac{n}{2}-i', j+\frac{n}{2}-j'} \\ (i^*, j^*) &= \text{ arg min}_{ij} a_{ij} \end{aligned} \tag{1}$$

again using two-dimensional indices and defining the *activity map* $A = \{a_{ij}\}$.

The weight update rule is identical to the original SOM rule:

$$\begin{aligned} \vec{\mu}^* &= (i^*, j^*) \\ \kappa_{ij}(t) &= \gamma\Big(\mu^*, \sigma(t), (i, j)^T\Big) \\ \vec{p}_{ij}(t+1) &= \vec{p}_{ij}(t) + \epsilon(t)\kappa_{ij}(t)(\vec{x}(t) - \vec{p}_{ij}), \end{aligned} \tag{2}$$

where the quantities $\sigma(t)$ and $\epsilon(t)$ have the following time dependence:

$$\begin{aligned} 0 < t < T_0 &: \sigma = \sigma_0, \, \epsilon = \epsilon_0 \\ T_0 < t < T_1 &: \sigma(t) = \sigma_0 e^{-\kappa(t-T_0)} \\ &\epsilon(t) = \epsilon_0 e^{-\kappa(t-T_0)} \\ t > T_1 &: \sigma = \sigma_\infty, \epsilon = \epsilon_\infty \end{aligned} \tag{3}$$

where $\kappa = \frac{\log(\sigma_\infty - \sigma_0)}{T_1 - T_0}$ is calculated such that $\sigma(T_1) = \sigma_\infty$.

The energy function minimized by this modified SOM model ("Heskes model") is

$$E = \text{min}_i \, a_i \tag{4}$$

using the definitions of eqn.(1) as demonstrated in [9]. It can be easily seen that this amounts to the original SOM rule for small values of $\sigma(t)$ as in this case the neighbourhood filter becomes a delta filter and thus the convolution has no effect.

The convolution in eqn.(1) can depass the boundaries of the distance map D, meaning that the indices in the sum can be come larger than n or smaller than 0. This problem has been long known in the computer vision literature when dealing with image convolutions [10], and is usually addressed by imposing particular boundary conditions. Traditional possibilities are *zero-padding* where all elements that do not fall into the image are treated as zero, or *periodic boundary conditions* where the image is considered a torus in both dimensions, and thus elements outside the image are taken from the opposite side of the image. Whatever boundary conditions are imposed, although they make convolution formally possible, they corrupt the integrity of the convolved image at its borders because they infer values for the original image that do not exist. Either, for zero-padding boundary conditions, convolution results are weak because of many adjacent zero values produced by the boundary conditions, or convolution results are uncorrelated with their neighbourhood as the opposite of the map influences the results. In this article, we propose a simple solution to this problem in the context of zero-padding conditions: we statically **re-weigh convolution results near the map borders** according to the part of the neighbourhood filter that falls outside the map. Without the re-weighting, activities would fall off towards the map borders because larger and larger parts of the neighbourhood filter are missing (or rather: applied to zeroes outside the image, thus not giving a contribution). As

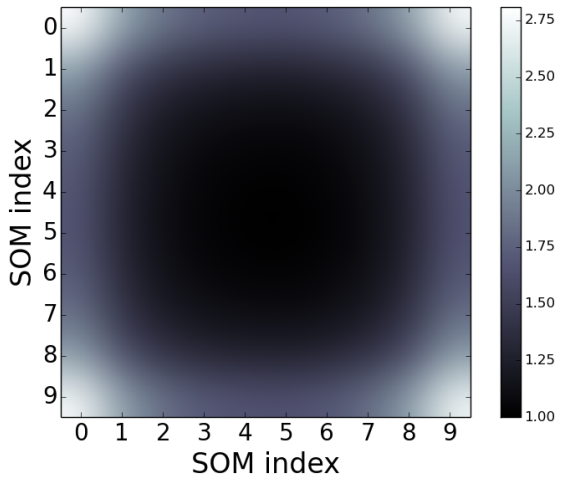

Fig. 2. Visualization of the inverted re-weighting map χ_{ij}^{-1} for a 10x10 map and a neighbourhood radius of $\sigma = 2$. It can be ob served that at the center of the map the values are equal to 1.0 corresponding to the fact that no correction for boundary effects is necessary there. In the corners, where the convolution filter has the lowest overlap with the distance map, correction factors grow strongly to compensate.

a consequence, units at the map border would effectively have no chance to ever become BMUs and thus the representational capabilities of the SOM would be impaired. The re-weighting is a multiplicative unit-wise operation:

$$\begin{aligned} \mathbb{1}_{ij} &\equiv 1 \\ \chi_{ij} &= (\mathbb{1} *_{zp} G)_{ij} \\ a_{ij} &\to \frac{a_{ij}}{\chi_{ij}} \end{aligned} \tag{5}$$

where we use the symbol $\mathbb{1}$ to denote a $n \times n$ map entirely composed of ones, which is convolved with the neighbourhood filter using zero-padding boundary conditions. The re-weighting map χ measures, at each point (i, j), which fraction of the mass of the applied neighbourhood filter lies within the map, see Fig. 2. Inserted into the definition of the activity map A in eqn.(1) and considering the energy function of eqn.(4), one perceives that the introduction of the re-weighting map does not depend on the model parameters (i.e., the prototypes $\vec{p}_i$) and thus does not affect the learning rule other than by a position-dependent constant factor. We will refer to this strategy of treating boundary conditions as "zero-padding with correction" (ZPC).

An alternative solution is to impose periodic boundary conditions for the convolution in eqn.(1). This solution is less complex and probably slightly faster. However, periodic boundary conditions for the neighbourhood filter imply the same type of boundary conditions for the neighbourhood function, which would amount to training a toroidal SOM. It is the purpose of this article to show that the Heskes model can indeed work for both types of boundary conditions, and that the proposed ZPC boundary conditions achieve significantly lower quantization errors (leaving all other parameters unchanged). This should be especially true for small map sizes, in which we are particularly interested as motivated previously.

B. Data

Two datasets are used in the experiments presented in this article. On the one hand, we rely on the well-known MNIST benchmark [12] for handwritten digit recognition that is a standard problem in machine learning. For our purposes, it is ideal for testing our implementations as it allows a visual inspection of the learned prototypes, facilitating the detection of implementation errors through obviously corrupted prototypes. However, in order to lend weight to our simulation results, we chose not to exclusively use a relatively clean, artificial dataset such as MNIST to validate our findings, but additionally a dataset coming from a real-world detection problem[1], the the Daimler Pedestrian Detection Benchmark[3]. This dataset is about pedestrian detection, i.e., the distinction of pedestrian images from background/non-pedestrian samples. Please see Fig. 1 for a visualization of the pedestrian-detection task.

Pedestrian detection is a difficult real-world problems which requires sophisticated preprocessing in order to be solved to any degree of precision. The preprocessing method applied to the image data is termed HOG (histogram of oriented gradients, see [2]) and represents a standard method in real-world visual object recognition. Using the terms of [2], the parameters of the HOG transform applied to cropped images downsampled to a size of 32×64 are: block size 16×16, cell size 8×8, 2×2 cells per block, 9 orientations bins and normalization enabled.

The MNIST dataset contains 60.000 training samples in 10 classes that are approximately equiprobable, as well as 10.000 samples in the test set. The pedestrian detection dataset contains 10.000 training samples and 19.148 test samples of dimension 756 (which arises from the particular parameters chosen for the HOG transform). It contains two classes, "pedestrian" and "background" which are very nearly equiprobable in both training and test sets. A visualization of the problem can be found in Fig. 1.

III. Experiments

A. Boundary conditions and quantization error

This simple experiment investigates the effects of different boundary conditions when training energy-based SOM models described in Sec. II-A. In particular, we compare the "zero-padding with correction" (ZPC) strategy of Sec. II-A with periodic boundary conditions (essentially a toroidal SOM). periodic boundary conditions impose more constraints on prototypes as they are imposed at all the boundaries of the SOM. In contrast, ZPC boundary conditions impose no constraints on prototypes close to the borders, leaving more freedom to model the data faithfully. The basic quantity we consider here is the so-called quantization error measure, which for a sample $\vec{x}$ and a set of prototypes $\{\vec{p}_i\}$ is defined as $q(\vec{x}, \{\vec{p}_i\}) = \min_i ||\vec{x} - \vec{p}_i||$, and which measures the capacity of a SOM to approximate its input.

The basic question we ask is: does the manner of treating boundary conditions show up in the quantization error of an energy-based SOM (see II-A). To investigate this, we train such a SOM, with both kinds of boundary conditions, for $T = 40.000$ iterations on the training set of the MNIST

[1] Available under www.gepperth.net/alexander/downloads/dataWSOM.tar.gz

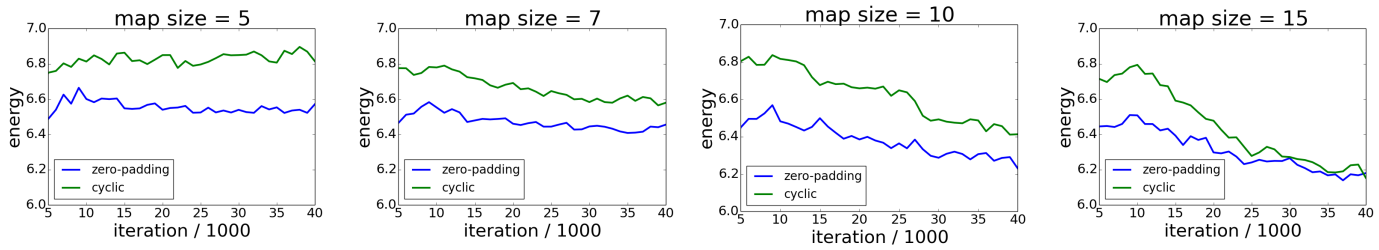

Fig. 3. Quantization errors for energy-based SOMs, comparing periodic and zero-padded+corrected boundary conditions, showing a clear difference between the two conditions especially for small map sizes. A smaller map size manifests itself by higher final quantization errors, and it is evident that larger SOMs suffer less from periodic boundary conditions than small ones.

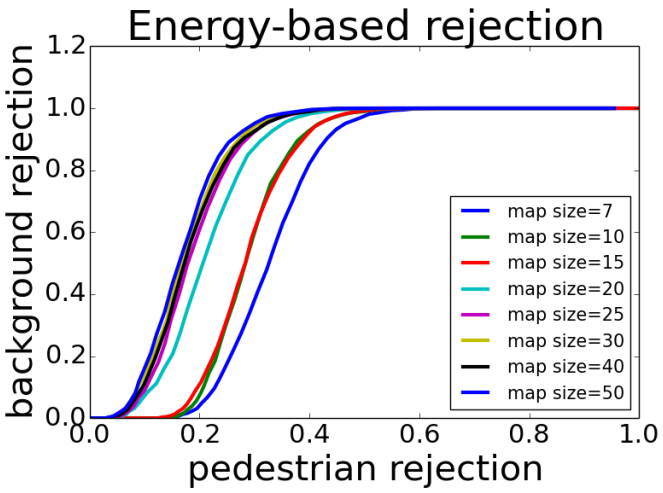

Fig. 4. Energy-SOM based outlier detection in the pedestrian detection task for various map sizes. Shown is the percentage of rejected pedestrians plotted against the percentage of rejected background samples. The energy-based outlier detection works rather well: for a map size of 50x50, for example, one may suppress 100% of background samples while only suppressing 30% of pedestrian samples.

problem described in Sec. II-B and measure its quantization error on the test set. Map size is varied in the interval $n = \{5, 7, 10, 15\}$ in order to asses the influence of the map size. The other parameters, in the terms of Sec. II-A are: $T_0 = 10000$, $T_1 = 30000$, $\sigma_0 = n/4$, $\sigma_\infty = 1.5$ (except for the case $n = 5$ where $\sigma_\infty = 1.0$), $\epsilon_0 = 0.1$, $\epsilon_\infty = 0.01$. We compare quantization errors on the MNIST test set over time for both boundary condition types, results are given in Fig. 3. As can be seen, the proposed ZPC boundary conditions work and generate SOMs with consistently lower quantization errors than periodic boundary conditions can reach, especially for small map sizes.

B. Energy-based rejection strategies

It is well known that SOM prototypes approximate the distribution of input vectors[1], and that thus the distance between an input vector and the prototype of the BMU can be used as a measure of a priori probability for this input vector. In other words, inputs the SOM has rarely or never seen will have a high distance d_{i^*} to the BMU prototype (with linear index i^*) and can thus be potentially identified. In this experiment, we wish to determine whether the same holds for the analogous quantity of energy-based SOM models, the BMU activity a_{i^*} (see Sec. II-A).

Concretely, we train several energy-based SOMs (with ZPC boundary conditions) of different map sizes n only on the "pedestrian" class of the pedestrian detection problem described in Sec. II-B and treat the "background" class as outliers of the pedestrian class that should be detected by imposing a threshold θ on the quantity a_{i^*}:

$$\text{class=} \begin{cases} \text{pedestrian} & \text{if } a_{i^*} < \theta \\ \text{background} & \text{else} \end{cases}$$

Training and test parameters are: $n = \{10, 15, 20, 30, 50\}$, $T = 70.000$, $T_0 = 20000$, $T_1 = 40000$, $\sigma_0 = n/4$, $\sigma_\infty = 0.05$, $\epsilon_0 = 0.1$, $\epsilon_\infty = 0.004$. As for this experiment a thorough convergence is necessary, more time was given to the energy-based SOM to converge, and a small asymptotic neighbourhood radius was used as this favored outlier detection. By logging the responses to test samples of a trained SOM, and by varying the threshold θ_D, one can obtain rejection curves as shown in Fig. 4, each point of which represents a different trade-off between retaining pedestrian and supressing background samples.

It can be seen from Fig. 4 that the rejection strategy based on the activity map works, and increasingly well so with larger map sizes. The interesting point here is that the basic quantity $a_{i^*} = \min_i a_i$ used for rejection here is nothing else but the instantaneous energy of the SOM given the current sample, please compare to eqn. (4). This means that, by optimizing the energy, one directly optimizes the quantity necessary for outlier detection, which is an elegant and intuitive approach.

C. Energy-based control strategies

As stated in the introduction, the behavior of the energy function can give valuable hints about controlling the time-dependent parameters $\sigma(t)$ and $\epsilon(t)$ in the SOM update rule 2. As any change in $\sigma(t)$ directly impacts, through the activity map a_{ik}, the energy function defined in (4), changes to σ should be adiabatic, i.e., small and effected only when the energy is stationary. This is the key point of the admittedly simplistic control strategy we adopt here, just to show what an energy function can do for SOM users: the detection of stationary energy states. This is effected by first exponentially smoothing the instantaneous energy values and subsequently computing a measure of how much the smoothed energy changed in a characteristic time interval defined by the constant

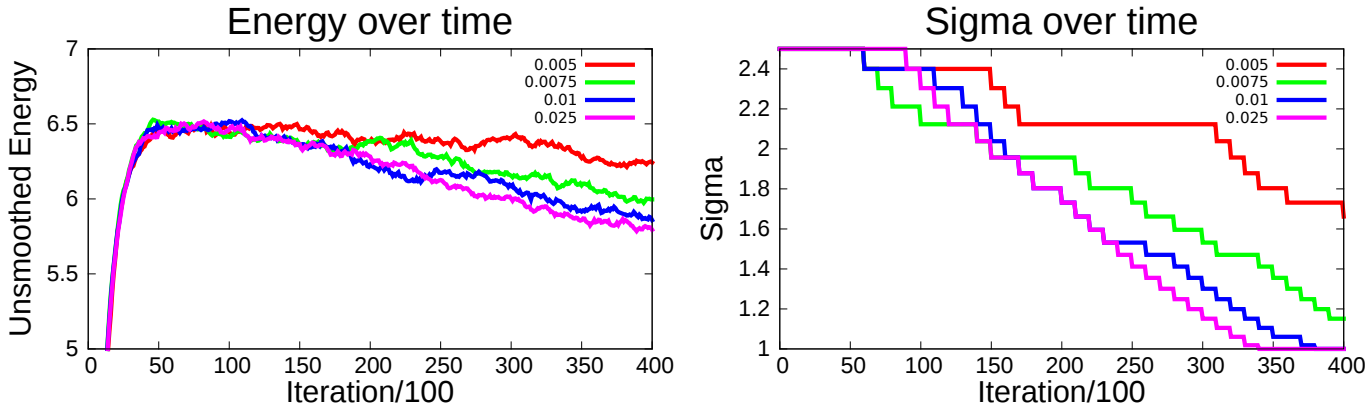

Fig. 5. Development of energy and neighbourhood radius over time using an automated control strategy for $\epsilon(t)$ and $\sigma(t)$ of the energy-based SOM. Experiments differ only in the value of the threshold θ_D that determines the upper limit on energy evolution that is allowed to consider the energy stationary. As expected, all values achieve an optimization of the energy function, but more restrictive values of θ_D result in slower convergence of SOM parameters, and in a slower decrease of the energy function. It may be noted that the evolution of σ is reminiscent of the exponential decay usually modeled when training SOMs, only here it arises automatically from the specified control law.

α:

$$\begin{aligned} E(t) &= \min_i a_i(t) \\ \hat{E}_\alpha(0) &= 0 \\ \hat{E}_\alpha(t) &= (1-\alpha^{-1})\hat{E}_\alpha(t-1) + \alpha^{-1}E(t) \\ D(t) &= 1 - \frac{\hat{E}_\alpha(t)}{\hat{E}_\alpha(t-\alpha)} \end{aligned} \tag{6}$$

and subsequently imposing a threshold on $D(t)$ to implement the following control strategy:

$$\sigma(t) = \begin{cases} \max\left(\sigma(t-1)\xi, \sigma_\infty\right) & \text{if } 0 < D < \theta_D \\ \sigma(t-1) & \text{else} \end{cases} \tag{7}$$

as well as an identical control strategy for $\epsilon(t)$. Here, the parameters $\theta_D \in [0,1]$ which determines how strict the definition of stationarity should be, and $\xi \in [0,1]$ which determines the decrease rate for ϵ and σ need to be specified. The time scale parameter α controls both the speed of exponential smoothing, as well as the time lag that is consider for the determination of stationarity. For the experiments of this section, we consider again the MNIST benchmark and train an energy-based SOM (with ZPC boundary conditions) using the control rule of eqn. (7), with fixed values of $\xi = 0.95$, $\alpha = 1000$ and variable values of $\theta_D = \{0.005, 0.0075, 0.01, 0.025\}$. The other SOM parameters are $n = 10$, $T = 40.000$, $T_0 = 10000$, $T_1 = 30000$, $\sigma_0 = n/4$, $\sigma_\infty = 1$, $\epsilon_0 = 0.1$, $\epsilon_\infty = 0.01$.

The control strategies differ thus only in the strictness of the definition of stationarity, and will therefore differ in their eagerness to trigger the next decrease step to ϵ and σ. As can, be seen in Fig. 5, all of these strategies are effective in reducing energy over time, although at different convergence speeds.

IV. Discussion

We presented an extension of the energy-based SOM model proposed in [9], characterized by the avoidance of periodic boundary conditions in the convolution operations associated with the model. We have shown that energy-based SOMs with ZPC boundary conditions achieve lower quantization errors than energy-based SOMs with periodic boundary conditions, especially for small map sizes (see Sec. III-A). We have furthermore shown that energy-based SOMs with ZPC boundary conditions can use the instantaneous energy value $E(t)$ as a criterion for rejecting outliers in a real-world visual pedestrian detection problem, and that the evolution of energy over time can be used for implementing automated control strategies for the time-dependent SOM parameters σ and ϵ that can remove the need to tune these time dependencies. Overall, we can state that ZPC boundary conditions make energy-based SOMs a tool that can be used as routinely as normal SOMS, without the need to resort to exotic and often undesirable periodic boundary condition that in any case reduce the approximation capacity of a SOM (not just an energy-based one).

A. Why ZPC works

The ZPC technique we introduced here is based on a simple static re-weighting of convolution results that form the activity map a_{ij}, see Sec. II-A. The reason why this approach works for SOMs (but not, in general, for arbitrary images) lies in the fact that we assume the temporal averages of all input-prototype distances d_i to be approximately homogeneous throughout the map. This is not the case for images, and thus the technique described here cannot be generalized to image processing. If our assumption holds true, then the temporal average activity a_{ij} of any SOM unit (after convolution) is indeed only dependent on the mass of the neighbourhood filter that falls inside the map, which is a quantity that depends only on that unit's position and on nothing else. The correction that fixes all map activities to homogeneous long-term values, here termed the re-weighting map χ can therefore, for fixed σ, even be precomputed, but needs to be re-calculated each time sigma changes. Obviously the assumption we make here is not rigorously provable from the energy function alone but would need to be ensured by an additional energy term governing long-term temporal averages. This would amount to computing long-term temporal averages for all SOM units by exponential smoothing and then using these measured averages for re-weighting purposes, thus constituting a generalization of the work presented here. As it is, the approach works extremely well for the presented datasets and a great number of other

problems.

V. Conclusion and future work

As stated in the previous section, it would be a safer approach for ZPC boundary conditions if the assumption of homogeneous temporal averages within the SOM were explicitly enforced by an additional energy term. Effectively, this would represent a kind of self-adaptation of the energy-based SOM to input statistics and could thus generalize the rather ad-hoc mechanisms of conscience and habituation[13] while putting them on the solid footing of gradient-based optimization procedure minimizing a composite energy function. More elaborate control strategies for time-dependent SOM parameters could very probably be devised (for example, there is a clear dependency between θ_D and α in Sec. III-C that can be used to eliminate one parameter) so as to render these strategies fully automatic, removing the need for any parameter tuning during SOM training.

References

[1] M. Cottrell, J.-C. Fort, and G. Pagès. Theoretical aspects of the som algorithm. *Neurocomputing*, 21(1):119–138, 1998.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[3] M. Enzweiler and D. Gavrila. Monocular pedestrian detection: Survey and experiments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(12):2179–2195, 2009.

[4] E. Erwin, K. Obermayer, and K. Schulten. Self-organizing maps: ordering, convergence properties and energy functions. *Biological cybernetics*, 67(1):47–55, 1992.

[5] A. Gepperth and C. Karaoguz. A bio-inspired incremental learning architecture for applied perceptual problems. *Cognitive Computation*, pages 1–11, 2016.

[6] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[7] T. Graepel, M. Burger, and K. Obermayer. Self-organizing maps: Generalizations and new optimization techniques. *Neurocomputing*, 21(13):173 – 190, 1998.

[8] T. Heskes. Energy functions for self-organizing maps. In *Kohonen maps*, pages 303–315. Elsevier, 1999.

[9] T. M. Heskes and B. Kappen. Error potentials for self-organization. In *Neural Networks, 1993., IEEE International Conference on*, pages 1219–1223. IEEE, 1993.

[10] B. Jähne. *Digital image processing*. Springer Verlag Berlin, Heidelberg, New York, 6 edition, 2005.

[11] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biol. Cybernet.*, 43:59–69, 1982.

[12] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. In S. Haykin and B. Kosko, editors, *Intelligent Signal Processing*, pages 306–351. IEEE Press.

[13] R. Rizzo and A. Chella. A comparison between habituation and conscience mechanism in self-organizing maps. *IEEE Transactions on neural networks*, 17(3):807–810, 2006.

[14] V. Tolat. An analysis of kohonen's self-organizing maps using a system of energy functions. *Biological Cybernetics*, 64(2):155–164, 1990.